I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12202

Peer Diversity, College Performance and Educational Choices

Arnaud Chevalier
Ingo E. Isphording
Elena Lisauskaite

MARCH 2019

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12202

Peer Diversity, College Performance and Educational Choices

Arnaud Chevalier
University of London and IZA

Ingo E. Isphording
IZA

Elena Lisauskaite
University of London

MARCH 2019

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Peer Diversity, College Performance and Educational Choices*

We study the effect of ethno-linguistic classroom composition in college on educational performance, educational choices and post-graduation migration in a setting of quasi-random assignment to undergraduate seminars at a British university. We focus on two core variables: the share of non-English-speaking students and the diversity within the group of non-English-speaking students with respect to their linguistic background. English-speaking students are largely unaffected by the ethno-linguistic classroom composition. Non-English-speaking students benefit from a larger diversity in their performance and increase their interaction with English-speaking students. Educational choices of non-English-speaking students become more similar to choices of English-speaking students in response to more diverse classes. Post-graduation, non-English students who have been assigned to higher shares of non-English students in the compulsory stage are more likely to leave the country. Our results imply that current levels of internationalisation do not impose a threat to native education. Avoiding segregation along ethnic lines is key in providing education for an internationalised studentship.

JEL Classification: I21, I24, J15

Keywords: higher education, diversity, peer effects, foreign students

Corresponding author:
Ingo E. Isphording
Institute of Labor Economics
Schaumburg-Lippe-Str.5- 9
53113 Bonn
Germany

E-mail: isphording@iza.org

* We would like to thank Benjamin Elsner, Jonas Radbruch, Ulf Zölitz, the participants of the Workshop on Higher Education at the University of Essex, the Workshop on the Economics of Education and Expectations at RHUL and the 3rd IZA Workshop on the Economics of Education. Excellent research assistance was provided by Maximilian Maehr and Youpeng Zhang. A. Chevalier would also like to thank the Nuffield foundation (grant EDU/42242) for partially funding this work.

1 Introduction

The tertiary education system has experienced extraordinary trends of fast-paced internalisation during recent decades. Between 2000 and 2013, the number of international students rose from 2.0 million to 4.1 million worldwide (UNESCO, 2018). The UK has become one of the largest recipient countries of foreign students. In 2016, it hosted more than 400,000 international students, representing 18% of the student population in the country. This development has led to an increasingly diverse student population, with a variety of ethnic and linguistic backgrounds. The emerging scientific and politic debate discusses benefits as well as potential detrimental effects of a diverse student body (Migration Advisory Committee, 2018). Advocates of internationalisation argue that increasing diversity benefits both native and foreign students, although critics raise concerns about potential negative spill-overs and native college flight. Empirical evidence on the effects of increasing numbers of foreign students is mostly restricted to primary and secondary education and points to ambiguous effects across different settings.[1] Tertiary education settings have received less attention to date.[2]

Against this background, we add to the understanding of how the ethnic composition of seminars (small-scale learning groups) can affect student performance and choices. We focus on two core variables: the share of students coming from a non-English language background (labelled non-English-speaking hereafter) and the ethno-linguistic diversity within the group of non-English-speaking students. Ethno-linguistic diversity among non-English-speaking students is expected to influence student performance and choices above and beyond the share as it changes incentives for English language use and assimilation. We relate these variables to student performance, course choices and post-graduation migration decisions. We base our analysis on the administrative records of economics students at a university in the London Metropolitan area. We further provide survey evidence on changes in social networks and language use as likely mechanisms. To alleviate concerns about selection, we rely on quasi-randomly assigned seminar classrooms.

Our institutional environment provides a fitting "laboratory setting" for our analysis for two reasons. First, the student body that we analyse is characterised by a high degree of ethno-linguistic diversity. Over the 2006 to 2016 period, we observe on average a share of 56

[1] The results range between moderate negative effects on native students' performance (Ballatore et al., 2018; Brunello and Rocco, 2013; Jensen and Rasmussen, 2011; Gould et al., 2009) to zero effects (Geay et al., 2013; Ohinata and van Ours, 2013).

[2] Anelli et al. (2017) show that higher shares of tertiary education students in introductory maths courses reduce natives' likelihood of moving into STEM majors. Braakman and McDonald (2018) point to ambiguous effects of diversity by the level of aggregation in UK universities.

percent of non-English-speaking students from 67 different non-English-speaking countries. The high degree of ethno-linguistic diversity ensures a sufficient common support of share of non-English-speaking students and the diversity to identify the causal effect of both variables. Second, students of the economics programme are quasi-randomly allocated to seminars during the compulsory stage of their studies, thus exposing them to exogenously-varying ethno-linguistic compositions in the classroom.

We describe four sets of results. First, the performance of English-speaking students is unaffected by the share of non-English-speaking students and the diversity of a classroom. Students from a non-English background display a slightly higher likelihood of failing a course when exposed to higher shares of non-English-speaking students, although this effect is mitigated by a higher ethno-linguistic diversity. In a more diverse classroom, the academic performance of non-English-speaking students improves, especially for low-achieving students.

Second, based on survey evidence, we show that students change their pattern of classroom interactions across ethnicities in response to higher diversity. When diversity increases, non-English-speaking students become more likely to interact with English-speaking students. We do not observe a similar effect of changed interaction patterns for English-speaking students. These first sets of results are consistent with a theoretical framework akin to Lazear (1999) in which incentives to invest in the majority language vary with the size of one's own language group. Ceteris paribus, a larger diversity among the non-English-speaking students is associated with a smaller own language group. Thus, diversity counteracts segregation into own-language learning groups and might have persistent effects through "English-learning by doing".

Third, the ethno-linguistic composition of seminars affects the future course choice of non-English-speaking students. When exposed to higher shares of non-English-speaking students in their early study seminars, non-English-speaking students choose more numerical classes and courses that are generally more popular among non-English-speaking students. We do not observe similar effects for English-speaking students. Early exposure to higher shares of non-English-speaking students thus appears to lead to segregation in subsequent stages of studies. This effect is again counteracted by more diverse first- and second-year seminars.

Finally, the results based on a voluntary alumni survey suggest that non-English-speaking students are affected in their decision to migrate after graduation. Being assigned to a larger share of non-English-speaking students in the compulsory stage increases the

likelihood of being abroad at the point of interview of the post-graduation survey.

Taken together, our findings suggest that 1) even at a high level of internationalisation, there are no negative effects of exposure to non-English-speaking students in small-class teaching on the learning of native students; 2) diversity improves the learning and integration of non-English-speaking students; and 3) the ethno-linguistic composition of seminars might affect return or onward migration decisions of foreign students.

With this paper, we add to the understanding of effects of the internationalisation of education. In general, existing studies have focused on the share of foreign students in primary and secondary school classrooms. Ballatore et al. (2018) exploit rules of classroom formation to identify a sizeable negative effect of additional immigrant students in Italian primary schools, which is substantially larger for foreign rather than native students. Other studies describe small negative to zero to even slight positive effects of higher shares of foreign students using variation between cohorts or classes of the same school (Gould et al., 2009; Geay et al., 2013; Ohinata and van Ours, 2013; Figlio and Özek, 2017; Frattini and Meschi, 2017; Diette and Oyelere, 2017), regions (Jensen and Rasmussen, 2011; McHenry, 2015; Hunt, forthcoming) or between countries (Brunello and Rocco, 2013). A related strand of the literature has examined native school and college choice responses to immigrant inflows (Betts and Fairlie, 2003; Machin and Murphy, 2018). Applying an empirical strategy based on within-school/across-cohorts variation in Dutch primary schools, Maestri (2017) finds a positive effect of more diverse classrooms on foreigners' language acquisition.

To date, tertiary education settings have received less attention. Anelli et al. (2017) show that higher shares of foreign peers in introductory maths courses reduces the probability of native students moving into STEM majors. Braakman and McDonald (2018) describe a relationship between diversity at the course level and student performance for three UK-wide cohorts and deal with potentially endogenous course choices by exploiting within-programme variation across courses. They complement their main analysis with an IV strategy relying on network effects among foreign students. Their results point to ambiguous effects of diversity, depending on the level of analysis.

Our paper makes three contributions to these strands of literature. First, we go beyond previous attempts to identify the effect of diversity relying on potentially selective variation between cohorts or regions; instead, we base our identification on a clean natural experiment relying on the quasi-random allocation of students to small-scale seminars. These seminars display relevant peer groups with meaningful social interactions. The as-good-as-random

assignment alleviates many concerns about potential confounders through the selection of students into seminars. Second, we provide insights into plausible mechanisms by surveying students in the field about ethnic interactions and language use. Third, we are able to provide evidence of the persistence of diversity on choices and performance during one's college career and post-graduation migration decisions.

Our results have implications for education practitioners. Even in an environment where non-English-speaking students represent more than half of the students, we do not find negative effects of their share on the performance and educational choices of English-speaking students. This supports current policies of pursuing greater internationalisation in higher education and should caution against forces asking for stricter admission policies discriminating by origin. Moreover, to favour the integration of non-English-speaking students, it is essential to diversify the ethno-linguistic composition of classrooms and avoid segregation by ethno-linguistic background. Our estimates suggest that this would display a no-cost Pareto-efficient policy.

Beyond the educational setting, our results may speak to the effect of diversity in work-related settings, where existing literature has not yet reached a consensus about the effects of diversity on productivity (Hoogendoorn and Van Praag, 2012; Kahane et al., 2013; Trax et al., 2015). Here, we provide complementary findings from "academic work groups" and show that higher levels of diversity indeed provide the potential to raise productivity and knowledge production.

2 Data and institutional setting

2.1 Institutional setting

We estimate the effect of the ethno-linguistic composition of seminars based on administrative data from an economics programme at a university in the London Metropolitan area. The university ranks among the top 30 universities worldwide with respect to the share of foreign students[3]. UK institutions of higher education are generally attractive to foreign students due to their recognised quality and the popularity of the English language. Moreover, UK higher education institutions have embraced the internationalisation of their students' body as the main source of income growth, since until recently the number of local students enrolled and their tuition fees were capped by the regulator. Importantly, these regulatory features mean that the inflow of foreign students has not been associated

[3]https://www.timeshighereducation.com/student/best-universities/international-student-table-2018-top-200-universities

with a crowding out of domestic students, whereby the composition of domestic students has not been altered by the inflow of foreign applicants (Machin and Murphy, 2018).

This institution is typical of the British higher education sector in terms of its organisation. It is publicly funded, selective, and tuition fees for home students are set at the maximum specified by the regulator.[4] Graduates from this institution earn the average graduate earnings five years after graduation (Britton et al., 2018). We focus on undergraduate students who are registered for any of the programmes offered by the economics department, either solely or in conjunction with other departments. Each cohort comprises about 200 students, split almost equally between English speakers and students from non-English-speaking countries.

Figure A1 describes the structure of undergraduate studies. In any given year, students take four teaching units. We focus on compulsory units that are taught over both Spring and Autumn terms, in either the first or second year of the undergraduate degree: "Principles of Economics", "Quantitative Methods I", "Quantitative Methods II", "Microeconomics", and "Macroeconomics". In addition, students can choose among optional courses that might encompass one or two terms. In their third year, course choice becomes entirely optional and students can choose from about 20 different courses. We regard third-year course choice as outcome variables in Section 4.4. Students need to obtain at least 40 out of 100 points to pass a course and a grade point average of 40 to pass a year.

Courses comprise weekly lectures taught by a faculty member and small-group seminars where students discuss their coursework assignments. Students are assigned on an unsystematic basis, which results in an as-good-as random allocation of peers (we provide respective balancing tests in Section 2.4, which empirically support the as-good-as random allocation). For each course, seminars represent between 30 and 50 percent of the instruction time. Teaching materials are developed by the course leader, and all seminar leaders receive the same instructions. Attendance at seminars is compulsory and monitored. Absenteeism may lead to exclusion from exams and in the case of non-EU students to visa revocation. Switching to a different group is prohibited. Seminars comprise 26 students on average, who meet for one hour per week for the full academic year. Seminars are taught by either the course leader or teaching assistants. Teaching assistants are usually PhD

[4]The regulator sets the maximum amount of annual tuition fees that can be charged to home and EU students, which is de facto the price charged by all institutions. Until 2015, the regulator also set the maximum number of home students enrolled, thus fixing the income that institutions could generate from local students' tuition fees. Institutions are free to take as many non-EU students and tuition fees for these students are uncapped. Over the period of study, the average fees for international students across British institutions was 3.5 times higher than the fees for local students (own calculation based on the Redding Survey of University Tuition Fees 2018-19, available at https://www.thecompleteuniversityguide.co.uk/university-tuition-fees/reddin-survey-of-university-tuition-fees/).

students in the economics programme and they are assigned to courses before the random assignment of students takes place. We later control for any effect of the seminar leader by including teacher fixed effects.

2.2 Sample description

Our sample comprises 8,505 student × seminar observations in five compulsory subjects (2,120 individual students in 332 seminars) during the period between 2006 and 2016. Students are matched to administrative registers by a unique identifier. For our estimations, we draw information about a student's contemporary performance in compulsory courses, course choices in the non-compulsory stage of their studies, post-graduation migration from a post-graduate survey as well as background characteristics (gender, age and nationality). We further administered a survey on mechanisms in a contemporary cohort.

Performance. Our first main outcome is the final grades received for compulsory courses in the first and second year of the studies. These grades are computed at the end of the year and include all of the coursework, mid-term exams and final exams. All coursework and mid-term exams are marked internally by the course leader and/or teaching assistants. Marking is undertaken anonymously. The final exam, which carries the highest weight towards the final grade, is marked independently by two internal graders and checked by an external moderator.

For our analysis of student performance, we construct three outcome variables. First, we standardise the final grade within each course per year. Courses with a final grade below 40 percent are defined as having been failed. Table 1 (Panel A) lists means and standard deviations of the performance variables. On average, 17% of students fail a course. At the other end of the grading scale, we observe whether a student finished the course *with honours*, i.e. with an average of 60% or above. About 60% of students in our sample have a final grade of 60 points or above.

Educational choices. We further analyse the effect of ethno-linguistic diversity on students' choices after the compulsory stage of their studies. In their third and final year of their studies, students choose four or more courses (depending on the length of the course, whether it is taught only in one semester or throughout the year) out of a set of about 20 different courses, irrespective of course or grade average pre-requirements. We operationalise a student's realised choice set by computing summary statistics of the chosen courses. First, we compute the share of numerical courses taken by a student. A course

is considered to be numerical if its content is mainly quantitative and the assessments comprise more calculations rather than essay-type questions. Second, we compute the average leave-me-out share of non-English-speaking students in the chosen courses over the full period of observation as a measure of course popularity among non-English speakers.[5] Third, we compute the average leave-me-out share of fails as a measure of difficulty of the realised choice set. These variables will serve as dependent variables to examine the effect of share and diversity on educational choices. In addition, we use the third-year data to compute own realised third-year grades. Since grades are based on the different courses that students have selected, and courses have different grade distributions, we standardise third-year grades by course. We use the average standardised grade as a measure of third-year performance.

Table 1 (Panel A) describes these variables based on third-year information. The self-selection of students into courses induces considerable variation over the realised choice sets. Choice sets differ in their average share of non-English-speaking students between 8 to 95 percent. The share of students failing in the chosen courses ranges from 0 to 46 percent. On average, 54% of students' choices in the final year are maths-intensive courses.

Post-graduation migration. To draw conclusions about a longer-term effect of ethnolinguistic seminar composition on students' post-graduation migration, we extract data from the Destinations of Leavers from Higher Education in the United Kingdom Survey (DLHE). The DLHE surveys recent graduates of each British institution six months after their graduation.[6] Graduates are contacted by e-mail, post and telephone. Being administered within the timeframe of post-graduation job search, it is not very informative about the students' career success. However, we observe whether a student has left the country.[7] We construct a binary indicator for being abroad at the time of interview. On average, 3% of former English-speaking students among survey respondents have left the country, compared with 16% of non-English-speaking students (Table 1, Panel A). These numbers might understate the true migration if migrants have lower response rates to the survey. We will later carefully examine the role of selective response when interpreting our results.

[5] We rely on the leave-me-out measures leaving out the current observation to ensure that own ethnicity and choice do not mechanically alter the outcome variables.

[6] The DLHE is organised by the Higher Education Statistical Agency (HESA). Prior to 2011/12, HESA only surveyed UK and EU domiciled graduates but since then the coverage has been expanded to all graduates, and the data used pertain to this latter period.

[7] Non-EU students can stay for up to four months after graduation, but can apply for visa extensions for further studies or working.

2.3 Ethno-linguistic composition of seminars

Language background by nationality. We do not have direct information on the language spoken by students. To classify the ethno-linguistic composition of seminars according to the language background of students, we therefore assign language by the nationality of a student. We classify students from an English-speaking country (where English is either the predominant or an official language) as English speakers. For non-English-speaking countries, we assign each student the predominant language of his/her nationality. While this is straightforward in most cases, we rely on a number of sources such as fact books and language encyclopaedias to determine the predominant language in the case of multilingual countries. Figure 1 and Table A1 in the appendix summarise the languages, related nationalities and number of speakers in our sample. Only 44 percent of our sample are classified as English speakers. The largest group of non-English-speaking students is Mandarin speakers (19 percent of the sample), followed by 5 percent Russian speakers, and 3 percent Italian speakers. Overall, our sample comprises students from 68 different language backgrounds.

Ethno-linguistic diversity. For each seminar in our sample, we compute the share of students originating from countries not having English as the predominant or official language, i.e. the non-English speakers. We then compute the diversity among the group of non-English speakers by seminar. Throughout the analysis, we will use leave-me-out measures of share and diversity, where the current individual observation is kept out of the computation. We define diversity by the Blau Diversity Index as

$$D = 1 - \sum_{k=1}^{K} p_k^2$$

where p_k is the fraction of language group k speakers among the non-English speakers, excluding the current observation. Intuitively, the Blau Index directly relates to the more commonly-known Herfindahl Index of concentration. It measures the probability that two randomly-drawn non-English-speaking students within a seminar group have the same language background.[8]

Theoretically-possible maximum levels of the Blau Index depend on group size and the respective maximum number of potentially-distinct languages in the seminar. We account for this property by dividing the Blau Index by its respective maximum value

[8]For very low numbers or shares of non-English students, both variables would have a discrete support. Given that only two seminars in our sample have five or fewer non-English-speaking students, we do not see discreteness of support as a threat to our empirical approach.

$$D_{max} = 1 - \frac{1}{n}$$

where n is the number of students in the group of non-English-speaking students in a seminar. This adjusted Blau Index has values ranging from 0, complete homogeneity, to 1, maximum possible heterogeneity of the non-English speakers group and is comparable across seminars. In Section 4.3, we test the robustness of our results against alternative measures of diversity based on nationality only, and by assigning predominant native languages to countries with English as the second official language.

Language skills and linguistic distance. International students are required to provide evidence of their level of English language proficiency, with the exception of students who graduated from international schools. The main two exams that are accepted are the Test of English as a Foreign Language (TOEFL) and the equivalent International English Language Testing System (IELTS). We only observe English proficiency scores for the 42 percent of non-English-speaking students who took the IELTS. To assess the potential heterogeneity of effects by own and average seminar language proficiency among the entire sample, we proxy language proficiency by the linguistic distance of a student's origin language towards English. We classify language backgrounds by their respective linguistic distance to English. Following Isphording and Otten (2014), we assign a distance measure from the Automatic Similarity Judgment Program (ASJP) database (Wichmann et al., 2018)[9]. This measure assesses the difference in pronunciations of a culturally-independent word list (the Swadesh list) and has been shown to be a good predictor of language skill differences between immigrants (Isphording and Otten, 2014).

In Figure 2, we assess how linguistic distance correlates with English proficiency for the subset of students with a recorded IELTS score. We indeed observe a strong negative correlation between language distance and IELTS score, which gives us confidence that linguistic distance to English is a good proxy for English proficiency among the population of incoming students.

2.4 Random assignment to seminars

The identification of a causal effect of ethno-linguistic seminar composition relies on the idiosyncratic nature of peer assignment to seminars. Course administrators in our setting are instructed to assign students to seminars on a purely unsystematic basis. Deviations from this unsystematic assignment should only be due to scheduling conflicts due to already-

[9]The ASJP database is publicly available at https://asjp.clld.org/

assigned seminars and lectures from parallel study programmes. Indeed, only information about study programmes is known to administrators when making the allocation, but no further student characteristics. We therefore assume the assignment of students to seminars to be as good as random.

Although we can rely on first-hand and in-depth institutional knowledge, we do not want to rule out the possibility that students, teachers or administrators circumvent the instructed unsystematic assignment out of reasons and through ways unknown to us. We therefore carefully test whether the observed patterns in the data are consistent with the assumed quasi-random assignment applying two different tests for quasi-randomisation.

First, we use the original data to simulate an artificial group assignment and compare the actual distribution of peer compositions against this simulated random distribution. For this purpose, we randomly assign students to artificial seminar IDs, accounting for differences in average seminar size in different courses (the average seminar size differs across courses between 16 and 37 students). We then compute the share of non-English speakers and the ethno-linguistic diversity in each of these artificial seminars. In Figure 3, the white bars show the distributions of share and diversity under simulated random assignment within courses per term and based on 1,000 permutations. The grey bars show the actual observed distribution. Both distributions are of similar shape. Permutation-based p-values of a Wilcoxon rank sum test for equality of both distributions do not reject the null at conventional significance levels ($p = .892$ for share and $p = .549$ for diversity).[10] The observed distributions therefore are a plausible outcome of the assumed quasi-random assignment to seminars within courses.

Second, we formally test whether observable pre-determined average seminar characteristics predict left-out individual characteristics. Under random assignment, no such systematic relationship should exist. We therefore regress individual-level characteristics on leave-out-shares/averages of the same characteristic (the mean value of the variable within the peer group, not accounting for the individual observation itself). We account for fixed effects for courses per year as the level where the randomisation takes place. Following Guryan et al. (2009) and Caeyers and Fafchamps (2016), we additionally control for the leave-out-share/average of the respective characteristic at the course/year level. This adjustment accounts for a mechanical negative correlation between own and peer character-

[10]To determine the permutation-based p-values, we compare single simulated draws under the null as large as the observed population with the overall simulated population based on all 1,000 draws using a rank sum test. The empirical p-values are determined as the share of simulated draws which generate a test statistic as or more extreme than the one resulting from the comparison between actual observed distribution (grey bars) with the simulation-based population (white bars) displayed in Figure 3.

istics that arises even under random assignment, as individuals cannot be their own peers. The results of this test are summarised in Table 2. Significant correlations no longer appear as soon as we control for the level of randomisation, for day/time fixed effects and study programme to account for potential deviations from random assignment due to scheduling conflicts.

Both implemented tests confirm that the observed seminar compositions in our data are consistent with the assumed random assignment of students to seminars. We therefore conclude that administrators, students, and teachers indeed do not infer in this process, and that we can maintain the necessary identification assumption of the quasi-random assignment of students to seminars throughout our analysis.

3 Empirical strategy

3.1 Empirical model

We measure the ethno-linguistic classroom composition by the share of non-English speakers and the ethno-linguistic diversity among the group of non-English speakers. We identify the causal effect of these two core variables of interest by exploiting the random assignment of students to seminars. The random assignment allows us to assume differences in ethno-linguistic composition to be unrelated to students' observed or unobserved characteristics. Irrespective of their own language background, a student can experience different shares of and diversity among non-English speakers in their seminar. A student is not able to self-select into seminars by their composition, or to select another course, since we are focusing on compulsory first- and second-year courses.

We estimate the effect of the share and diversity of non-English speakers via

$$y_{ics} = \beta_1 share_{non-English,cs} + \beta_2 D_{cs} + X_i'\gamma + Z_{cs}'\delta + \theta_{\mathbf{c}} + \epsilon_{ics}. \tag{1}$$

Here, y_{isc} denotes outcomes for student i, taking course c (the subscript c denotes a specific course in a specific year) and assigned to seminar s. The main variables of interest, $share_{non-English,cs}$ and D_{cs}, are the leave-me-out share of non-English-speaking students and ethno-linguistic diversity, respectively, assigned at the level of the seminar s. We additionally control for individual student characteristics (X_i), age, gender, whether they are non-English speakers and the distance of the language of their country of birth to English. Z_{cs} is a vector of seminar-level characteristics including seminar leader fixed effects, size of the seminar and the leave-me-out share of observable characteristics. We

additionally include course× year fixed effects to capture any unobservable characteristics that would be shared by all students attending a certain course in a specific year. This is also the level at which the randomisation takes place. ϵ_{ics} is the error term. We cluster standard errors at the seminar level relying on 332 clusters. We later corroborate our inference by clustering on larger levels of aggregation and using empirical p-values based on permutation exercises in Section 4.3.

3.2 Identifying assumption and variation

We rely on a setting with as-good-as random assignment of students to seminars. While this random assignment alleviates concerns about the self-selection of students into peer groups, our identification still relies on the assumption that peer ethnicity is, conditional on observable dimensions, unrelated to unobservable peer characteristics that affect the outcomes. We later provide empirical support for the validity of this assumption by examining the coefficient stability with respect to controlling for observable peer characteristics like gender, age, residence and prior achievement.

To separately identify the effects of share and diversity under this identification assumption, we have to observe sufficient co-variation in both variables. Figure 4 provides a schematic description of the variation in our core variables. Seminars can differ in the share of non-English-speaking students (black symbols). Conditional on a specific share of non-English speakers, seminars can differ in their level of diversity among the group of non-English speakers. Comparing the hypothetical seminar B and C, it is easy to see that with a given share of non-English-speaking students, there can be variation in the diversity, here between 0 and 1. Moving from a group of non-English-speaking students that is fully homogeneous (seminar B, comprising speakers from one single group) to a seminar where the group of non-English speakers comprises many different language groups (seminar C) increases the diversity while keeping the share constant.

Figure 5 displays the common support in both the share of non-English speakers and the Blau Index in our raw data (left panel). We observe considerable variation in the share of non-English speakers at the seminar level, ranging from about 20 to over 80 percent. For each given share, there is also considerable variation in the diversity measure.

The right panel describes the respective variation in residuals after we account for fixed effects at the course×year level and control for individual characteristics and seminar characteristics. This is the relevant variation that is used in the quasi-experimental setup. Table A2 in the appendix summarises the residual variation left in key variables after we

control for the fixed effects, student and seminar characteristics according to equation 1. The standard deviation of the share of non-English speakers reduces between raw measures and residuals from 0.14 (absolute) to 0.09 (residuals), and from 0.10 (absolute) to 0.07 (residuals) for the Blau Index. Running a regression of the share of non-English-speaking students on the diversity and controlling for fixed effects according to equation (1) yields an insignificant coefficient of diversity of $\beta = 0.088[0.062]$. This low partial correlation allows separately identifying effects of the share of non-English speakers and the diversity of the non-English speaker group. For the remainder of the paper, we will refer to the standard deviations in residuals when describing effect sizes.

4 Results

We first present our main results on the effect of ethno-linguistic classroom composition on the contemporary performance of students in Section 4.1. We use survey evidence to describe the potential mechanisms behind the main results in Section 4.2. We test for the sensitivity of our results to different definitions of diversity and corroborate our inference with non-parametric permutation exercises in Section 4.3. We then turn to evidence on longer-term effects on third-year choices in Section 4.4 and post-graduation migration decisions in Section 4.5.

4.1 Performance

We start the discussion of the results with the effects of ethno-linguistic seminar composition on the contemporary performance of students. Both the share of non-English speakers and their diversity shape the learning environment of students. The share of non-English-speaking students will reduce the average English proficiency level in the classroom for both English and non-English speakers and might have negative spill-overs. However, higher diversity is expected to increase incentives to engage in English conversation for those of non-English language background and might lead to positive performance effects.

The results of estimating equation (1) on contemporary grades, and indicators for failing and receiving honours are summarised in Table 3. The table is organised in three panels describing the average results for all student-seminar observations (upper panel), English speakers only (middle panel), and students from a non-English language background (lower panel). For the full sample, the share of non-English speakers does not appear to affect contemporary grades (column 1). Nonetheless, we observe a marginally significant positive effect of diversity. A one standard deviation higher diversity increases grades by 1.5% of a

standard deviation. For the probability of failing a course, we observe a weakly significant relationship at the 10 percent level with a higher share of non-English-speaking students (column 2). Increasing the share of non-English-speaking students by 10 percentage points increases the probability of failing a course by 0.75 percentage points from a sample mean of 16.7%. As failing primarily affects lower-achieving students, we interpret this as evidence of a small negative effect of the share of non-English-speaking students on the academic achievements of low-performing students. Higher diversity counterweights this negative effect of the share. Holding constant the share of non-English-speaking students, a higher diversity by one standard deviation ($sd = 0.07$) reduces the likelihood of failing by 0.8 percentage points. We do not find effects on finishing a course with honours.

The average effects on the total sample mask a significant degree of heterogeneity. The effect is primarily driven by non-English-speaking students. For English-speaking students, coefficients for both the share of non-English-speaking students and diversity are small and remain insignificant (middle panel). English-speaking students are therefore unaffected by both the number and the diversity of surrounding non-English-speaking students.

The lower panel summarises the effect on non-English-speaking students. Non-English-speaking students strongly benefit from being assigned to a more diverse seminar. Increasing diversity by one standard deviation increases grades by 3.1%. The detrimental effect of the share of non-English on the probability of failing is larger than in the total sample, but remains insignificant, potentially due to the smaller sample size. The effect is again counteracted by a strong positive effect of higher diversity on performance. A higher diversity by one standard deviation reduces the probability of failing by 1.1 percentage points. We find an effect on the probability of achieving an honour-level grade. The positive effects of diversity therefore appear to be concentrated among the lower-achieving non-English-speaking students.

Taken together, these results suggest that a higher share of non-English speakers has a small, statistically marginally significant negative effect on the probability of passing the course, and no effects on grade performance. This is consistent with the moderate to zero effects of foreign students on native performance found in previous studies (Brunello and Rocco, 2013; Jensen and Rasmussen, 2011; Gould et al., 2009; Geay et al., 2013; Ohinata and van Ours, 2013). Thus, it appears that the institution studied by us is not out of line with other education settings regarding the effect of non-native students. This effect is counteracted by a strong positive effect of having a more linguistically-diverse environment.

11

4.2 Mechanisms

We now explore potential mechanisms for the contemporary effects of ethno-linguistic classroom composition on student performance. We observed that the performance of English-speaking students is largely unaffected by larger shares of non-English-speaking students and higher diversity. We therefore focus on language-based mechanisms as the most plausible candidate.

Our findings support an adaption of a model akin to Lazear (1999) where the value of investing in using the majority language is greater to a member of a small linguistic minority than one of a large minority. In the model by Lazear (1999), individuals learn languages to benefit from a larger pool of trading partners. Thus, the value of language adoption is greater to small minorities than large minorities and diversity-increasing policies may increase welfare through increasing incentives for assimilation.

Such a model can be easily adapted to the classroom. Instead of trading, students engage in co-learning. A student from a minority broadens the pool of potential learning partners if she engages in English communication, albeit potentially at some cost if she is not a native speaker. The smaller the pool of potential same-language learning partners, the higher the incentives to engage in English communication. This intuition directly maps into a theoretically-positive effect of diversity on classroom integration and, in turn, the quality of the learning environment, conditional on the share of non-English-speaking students.

The administrative data of the main analysis does not contain information on such a mechanism. We therefore collected additional survey data on the most recent cohort of students attending the same courses as those investigated in the main analysis. Specifically, we used an in-class written questionnaire asking students about the frequency of interactions with students by language background (own language group and English-speaking students), their English use and proficiency, and perceived quality of English in the classroom. Additionally, we asked about course-specific study hours.

We obtained data from 538 student×seminar observations, 222 of them with a non-English background. The overall response rate of the survey was 51 percent. Non-response

[11] In Appendix Table A3, we examine the sensitivity of the results against a measure of own and peer ability. For most foreign students, the records available to us do not contain a measure of pre-university academic ability. Instead, we approximate ability by test scores in other courses and for every student i in seminar j we compute the students' leave-out GPA as $GPA_{ij} = \frac{1}{K}\sum Grade_{i,k\neq j}$. Our observed patterns are robust against this coarse control for ability. We still report no significant effects of diversity for English speakers, and positive effects on test scores, especially for low performers, among non-English-speaking students. However, we acknowledge that this measure will display a bad control in case of spill-overs between subsequent courses.

is unrelated to the seminar linguistic composition.[12] We merged survey responses to seminar composition variables of the share of non-English-speaking students and diversity. The survey questions were mainly asked on 1-5 scales regarding the frequency of interactions or quality of language proficiency, which we standardise for our empirical exercise. The questionnaire questions are displayed in Table A7 in the appendix.

Table 4 summarises the results of estimating equation 1 for different items on language use and interaction by ethno-linguistic background as outcomes. The results strongly support language use as an important mechanism behind the observed performance results. In seminars with a larger share of non-English-speaking students, they are substantially less likely to interact with English-speaking students (lower panel, columns 1) and more likely to interact among each other (column 2). We do not observe similar effects for the English-speaking students (middle panel). This segregating effect of a larger pool of non-English-speaking students is mitigated by higher diversity among the non-English-speaking students. Facing a more diverse classroom and holding the share of non-English-speaking students constant increases the interaction of non-English-speaking students with their English-speaking fellow students. This pattern is in line with the idea sketched above of incentives to engage in interactions with the majority, which increase with the level of diversity. Effects on direct measures on perceived comfort in English use and the quality of English in the classroom remain insignificant for non-English-speaking students. English-speaking students perceive a lower quality of English spoken in the classroom with a higher share of non-English-speaking students.

Finally, we observe a decrease in self-reported study hours of non-English-speaking students (but not of English-speaking students) in seminars with a higher share of non-English-speaking students. A higher share of non-English-speaking students by 10 percentage points is associated with a reduction by 1.1 hours (from a mean of 5.8 hours). This reduction in study hours is a likely mechanism explaining the increase in failing rates in response to a higher share of non-English-speaking students discussed in the previous section.

Taken together, the survey evidence supports the idea that different incentives to engage in interactions with English-speaking students are a main mechanism to explain the effects of diversity on performance.

[12] Regressing a binary indicator of non-response on the share of non-English-speaking students and diversity using equation 1 yields small and insignificant coefficients of $\beta_{share} = 0.15[0.22]$ and $\beta_{diversity} = 0.02[0.31]$.

4.3 Robustness checks

Alternative definitions of diversity. In our main specifications, we define diversity along the lines of language groups. As such, foreign-born students from English-speaking countries are defined as English-speaking students. This definition already anticipates language being a main mechanism in terms of how diversity affects student performance. Nonetheless, diversity could be defined along related but different dimensions.

In Table A5, we replicate the main findings of Table 3, column (2), using alternative definitions of diversity and exposure to foreign students. Specifically, we test two alternative definitions. Column (1) lists the baseline results. In column (2), we deviate from the main specification by assigning the predominant language spoken to countries with English as an official language.[13] While the general pattern remains, the positive effect of diversity is substantially reduced. One potential explanation is that we now include students who speak English very well (as it is an official language in their country of birth) as part of the non-English-speaking population. These students are potentially less affected by a larger diversity. In column (3), we define nativity and diversity solely on the basis of nationality. Accordingly, we compare UK students to non-UK students and disregard the language dimension; thus, estimates would mix the effect of language and of culture. As before, the pattern of estimated coefficients remains similar to our main specification, but loses precision. Taken together, patterns of results are fairly stable across different definitions of diversity and nativity.

Robustness of inference. In our main specifications, we allow for clustering of error terms at the seminar level. In Table A9, we examine the robustness of our inference towards different inference corrections. For reasons of clarity of exposition, we focus on our preferred specification of Table 3, column (4), which is replicated in column (1). We subsequently alter the level of inference correction. Column (2) lists results for assuming i.i.d. error terms. Column (3) applies simple robust standard errors. In the remaining specifications, we adjust the level of clustering to the course×term level (column 4) and the term level (column 5). Standard errors of the share parameter increase with higher levels and smaller numbers of clusters. Standard errors of the diversity parameter appear to be insensitive. Our conclusions are thus unaltered by the choice of inference correction.

Finally, we assess the robustness of inference by using non-parametric permutation tests. For this purpose, we randomly assign students within courses to placebo seminar IDs and

[13] Gambia, Kenya, Nigeria, South Africa, Trinidad & Tobago, Uganda

re-run the analysis. We repeat this procedure 2,000 times. Distributions of the resulting simulated coefficients in relation to the originally-estimated parameter are summarised in Figure A2 in the appendix, focusing on the main results of Table 3, lower panel. Implied empirical p-values confirm the parametric significance levels.

Correlated peer characteristics. The quasi-random assignment of students to seminars alleviates common empirical issues related to the selection of students into seminars. Nonetheless, the share of non-English-speaking students might still be correlated with further dimensions of peer characteristics that have spill-overs on choices and performance by themselves, e.g. gender and ability. We carefully control for leave-me-out averages and shares of those variables that we observe: age, gender, and the linguistic distance between country of birth and English. We further approximate peer ability through peer achievement. In the following, we evaluate the stability of coefficients between these different specifications.

Table A6 in the appendix summarises coefficients for the share of non-English-speaking students and the diversity among them for four different specifications including only individual characteristics (1), additional seminar characteristics (2), and repeat those specifications by additionally controlling for own (3) and peer achievement (4). The comparison of specifications with and without seminar characteristics (1 vs. 2, and 3 vs. 4) is informative about the role of seminar characteristics as potential confounders. Estimated coefficients of diversity are fairly stable when controlling for additional peer characteristics. This is not surprising: peer characteristics other than the diversity appear not to play a role in explaining the outcome, whereby the according R^2 values barely change between specifications. Thus, we are unable to construct formal parameter bounds following Altonji et al. (2005) and Oster (2017) but conclude that all peer observations that are observable in our data appear do not interact with the observed peer effects by linguistic background.

While this robustness check does not hint at confounding unobserved peer characteristics, we cannot rule out that observed effects are picking up variation in peer characteristics that are uncorrelated with observed peer controls. This issue of potentially confounding but unobserved peer characteristics is common to any peer effects study that relies on natural variation in peers. Confounding peer characteristics cannot be separated from a person, and therefore cannot be (quasi-)experimentally stimulated. However, one might argue that owing to this inseparability, estimated effects are the relevant policy parameter. Finally, we argue that while shares and averages in further unobserved dimensions might be correlated with the share of non-English-speaking students, this is much less likely to affect our main

parameter of interest, namely the effect of diversity in the group of non-English speakers.

The role of Chinese students. Diversity and share of non-English speakers are driven to a large degree by the share of Chinese Mandarin speakers, representing 19 percent of our sample. We therefore test the robustness of our results with respect to controlling for the share of Mandarin speakers. We further examine heterogeneity in effects between Mandarin speakers and other non-English-speaking students.

Regressing the share of Mandarin speakers on the share of non-English speakers and the diversity and controlling for fixed effects according to equation (1) yields coefficients of $\beta_{share} = 0.28[0.024]$ and $\beta_{diversity} = -0.61[0.028]$ (Table A8, column 1). Column (2) displays the baseline results from Table 3. Despite this strong correlation between the share of Mandarin speakers and our variables of interest, additionally controlling for the number or the share of Mandarin speakers in the seminar preserves the general pattern of the main results (columns 3 and 4), yet the coefficients become insignificant. Splitting the sample between Mandarin speakers and other non-English-speaking students (columns 5 and 6) again yields patterns similar to our main results for both groups. Again, coefficients for diversity are less precisely estimated and insignificant, albeit likely due to sample size issues. We conclude from the stability of patterns that although Chinese play an important role in generating our results, the underlying mechanisms are to some degree independent of their presence.

4.4 Course choices

We now turn to the effects of ethno-linguistic seminar composition on final-year course choice and performance. Early seminar composition might have longer-run effects on future grades if peer interactions or learning behaviour acquired in the compulsory stage continue to influence educational attainment. There are at least three reasons why we expect educational choices to be affected. First, the initial effect on grades affects students' perceived academic ability, which might sway their choices towards more or less demanding courses. In particular, non-English speakers might reinforce their perceived comparative advantage in more quantitative courses, while English speakers similarly might perceive their comparative advantage in English as being even greater. Second, students exposed to more individuals from other ethnicities change their patterns of interaction, which might change their attitudes towards these ethnicities (Boisjoly et al., 2006; Carrell et al., forthcoming), thus making them more willing to interact with the same ethnicities in future courses. Third, having experienced a linguistically-dissonant learning environment in their

compulsory stage, students might opt for courses with a more quantitative curriculum where verbal communication plays a lesser role (Anelli et al., 2017).

We focus on three indicators describing a students' realised choice set among third-year non-compulsory courses: share of numerical courses, popularity among non-English speakers (measured as the historical share of English-speaking students) and difficulty (measured as the historical share of fails in the chosen courses). Note that these outcomes only vary at the student level, while the treatment measured by the share and diversity of non-English-speaking students varies at the student×seminar level.

Table 5 reports estimates of equation (1) for final-year outcomes. It is organised into three panels separately for all students, English speakers and non-English-speaking students. The results are primarily driven by non-English-speaking students: all parameters are insignificant for English-speaking students.

For non-English-speaking students, we observe a performance gain in third-year grades in response to having met a larger share of non-English-speakers in the first two years. This positive effect of the share of non-English speakers on third-year performance is at odds with a weak negative effect of the share of non-English-speaking students on low-performing non-English-speaking students for contemporary performance described in Section 4.1. This difference in effects implies different mechanisms affecting performance in the short and medium run. One potential explanation may be found in adjusted learning technology, i.e. the substitution of classroom learning with self-study efforts, producing lasting medium-run effects.

Higher shares of non-English-speaking peers in the compulsory stage further affect the likelihood of non-English-speakers choosing classes that are in general more popular among non-English-speaking students, as well as selecting a greater share of numerical courses. Both results imply a higher segregation and clustering among non-English-speaking students when initially exposed to more non-English-speaking peers. A higher share of non-English-speaking students by 10 percentage points in the compulsory-stage seminars increases the observed share of non-English peers in the realised third-year choice set by 0.58 percentage points.

A higher language diversity in compulsory-stage seminars mitigates this segregation. Being assigned to a seminar with a higher diversity by one standard deviation ($sd = 0.07$) reduces the share of non-English peers in the third year by 0.35 percentage points. This is in line with the survey evidence whereby more diverse compulsory-stage seminars broaden the networks of non-English-speaking students to English-speaking students. The results

are further consistent with the idea that diversity alters the preferences of non-English-speaking students for mixing with English-speaking students and increasing the similarities in module preferences.[14]

4.5 Post-graduation migration

We now turn to the analysis of longer-term effects on post-graduation migration. While foreign students in general cover costs of education through substantial fees, host countries gain from graduates remaining in the country to work. The migration advisory committee has stressed in its current report that international students have been able to "*fill shortages e.g. in STEM jobs, tying in with the government's Industrial Strategy.*" (Migration Advisory Committee, 2018). Therefore, the report suggests altering current visa regulations to ease the transition of foreign graduates into the labour market. However, what are the determinants that make students want to stay in the UK?

In Table 6, we examine the role of ethnic classroom composition during the compulsory stage of their study as a potential candidate. These results are based on the DHLE survey, so first it is important to check whether survey participation is correlated with the initial seminar allocation. Compulsory-stage seminar characteristics do not affect the post-graduation survey participation of non-English-speaking students. The low response rate to the graduate survey of non-English speakers is not systematically related to the ethno-linguistic seminar composition. English speakers are less likely to respond if they are exposed to more diverse seminars, although the effect is small. Being exposed to a higher diversity by one standard deviation reduces the response probability by 2.1% from a mean of 40.5 percentage points. Nonetheless, this effect of treatment on the response rate cautions interpreting the effect on English speakers as causal.

The results on migration decisions differ between English- and non-English-speaking students. Non-English-speaking students who are exposed to more non-English peers in their compulsory stage are more likely to have left the country at the point of the survey. A higher share of non-English peers in a compulsory-stage seminar by 10 percentage points increases the respective probability by 3.4%. This effect is in line with having fewer opportunities to build English-based networks in case of non-English-dominated seminars. We cautiously interpret these results as suggestive evidence of an effect of exposure to non-English-speaking students on their return or onward migration. English-speaking students are unaffected in their migration decisions by the ethno-linguistic composition of

[14]We report in Appendix Table A4 that controlling for own and peer ability does not alter these conclusions.

compulsory-stage tutorials.

5 Conclusion

Using data from a UK higher education institution that quasi-randomly allocates students to small classes, we do not find strong evidence of a negative effect of having a larger share of non-English-speaking students. In particular, English-speaking students are unaffected by the ethno-linguistic composition of seminars. Non-English-speaking students benefit from higher ethno-linguistic diversity in terms of their performance. Higher diversity makes the educational choices of non-English-speakers more similar to those of English-speaking students. Survey evidence implies that diversity augments the interaction among English and non-English-speaking students.

Our results are informative for the design of classroom assignment processes. Finding no negative spill-overs of either the share or diversity of non-English-speakers on English-speaking students, our results suggest that strategically avoiding segregation into classrooms may enhance the performance and integration of foreign students. Increased diversity improves grades, can be achieved at no cost and since it has no negative effect on any group, it is Pareto efficient. Our recommendation is that assignment should be made by stratified randomisation to seminars, where students are randomised within their own language groups.

The group-work-focused learning environment in the seminars may allow for generalisations to other settings of team production involving cognitive tasks. Evidence of the effect of diversity in production settings is scarce, restricted to either quasi-experiments based on sports data (Kahane et al., 2013), lab evidence (Hoogendoorn and Van Praag, 2012) or descriptive evidence from observational data (Trax et al., 2015). Here, we add causal field evidence from a setting sharing many features of collaborative environments, which are now standard in many workplaces.

More generally, the effects of diversity on economic and social outcomes appear to differ by the level of aggregation and results are inconclusive so far. Alesina and Ferrara (2005) propose a model that allows for negative effects of diversity on public good provision and positive effects on productivity. The majority of the literature so far has focused on the former negative effects, with recent evidence by Algan et al. (2016) demonstrating the negative effect of diversity on social cohesion in housing blocks in France. The latter positive effect on productivity has only recently gained attention by linking higher productivity (income) to birth place diversity (Ottaviano and Peri, 2006; Ager and Brückner, 2013).

Against this broader literature on diversity, our results are informative about the positive effects of diversity on productivity on a much smaller level of peer groups with strong and meaningful social interactions.

References

Abadie, A., Athey, S., Imbens, G. and Wooldridge, J. (2017). When Should You Adjust Standard Errors for Clustering? *NBER Working Paper* 24003.

Ager, P. and Brückner, M. (2013). Cultural diversity and economic growth: Evidence from the us during the age of mass migration. *European Economic Review* 64: 76–97.

Alesina, A. and Ferrara, E. L. (2005). Ethnic Diversity and Economic Performance. *Journal of Economic Literature* 43: 762–800.

Alesina, A., Harnoss, J. and Rapoport, H. (2016). Birthplace diversity and economic prosperity. *Journal of Economic Growth* 21: 101–138.

Alesina, A. and La Ferrara, E. (2002). Who trusts others? *Journal of Public Economics* 85: 207–234.

Algan, Y., Hemet, C. and Laitin, D. D. (2016). The Social Effects of Ethnic Diversity at the Local Level: A Natural Experiment with Exogenous Residential Allocation. *Journal of Political Economy* 124: 696–733.

Altonji, J. G., Elder, T. E. and Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economy* 113(1): 151–184.

Anelli, M. and Peri, G. (2017). The Effects of High School Peers' Gender on College Major, College Performance and Income. *Economic Journal, Forthcoming* .

Anelli, M., Shih, K. and Williams, K. (2017). Foreign Peer Effects and STEM Major Choice. *IZA Discussion Paper Series* 10743.

Ballatore, R. M., Fort, M. and Ichino, A. (2018). Tower of Babel in the Classroom: Immigrants and Natives in Italian Schools. *Journal of Labor Economics* 36.

Betts, J. R. and Fairlie, R. W. (2003). Does immigration induce 'native flight' from public schools into private schools? *Journal of Public Economics* 87: 987–1012.

Boisjoly, J., Duncan, G., Kremer, M., Levy, D. and Eccles, J. (2006). Empathy or Antipathy? The Impact of Diversity. *American Economic Review* 96: 1890–1905.

Braakman, N. and McDonald, S. (2018). The Impact of Student Diversity on Student Outcomes at University and Beyond - Evidence From English Administrative Data. *Unpublished manuscript* .

Britton, J., Belfield, C., Erve, L. van der, Vignoles, A., Walker, I. and Dearden, L. (2018). Estimating the Returns to College Degrees Using Linked Administrative Data. *Institute for Fiscal Studies Research Report* .

Brunello, G. and Rocco, L. (2013). The Effect of Immigration on the School Performance of Natives: Cross Country Evidence Using PISA Test Scores. *Economics of Education Review* 32: 234–246.

Caeyers, B. and Fafchamps, M. (2016). Exclusion Bias in the Estimation of Peer Effects. NBER Working Papers 22565, National Bureau of Economic Research, Inc.

Carrell, S. E., Hoekstra, M. and West, J. E. (forthcoming). The impact of college diversity on behavior toward minorities. *American Economic Journal: Economic Policy* .

Diette, T. M. and Oyelere, R. U. (2017). Do limited english students jeopardize the education of other students? lessons from the north carolina public school system. *Education Economics* 25:5: 446–461.

Feld, J. and Zölitz, U. (2017). Understanding Peer Effects: On the Nature, Estimation and Channels of Peer Effects. *Journal of Labor Economics* 35: 387–428.

Figlio, D. N. and Özek, U. (2017). Unwelcome Guests? The Effects of Refugees on the Educational Outcomes of Incumbent Students. NBER Working Papers 23661, National Bureau of Economic Research, Inc.

Frattini, T. and Meschi, E. (2017). The Effect of Immigrant Peers in Vocational Schools. *IZA Discussion Paper Series* 11027.

Geay, C., McNally, S. and Telhaj, S. (2013). Non-Native Speakers of English in the Classroom: What Are the Effects on Pupil Performance? *The Economic Journal* 123: 281–307.

Gould, E. D., Lavy, V. and Paserman, M. D. (2009). Does Immigration Affect the Long-Term Educational Outcomes of Natives? Quasi Experimental Evidence. *The Economic Journal* 119: 1243–1269.

Guryan, J., Kroft, K. and Notowidigdo, M. J. (2009). Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments. *American Economic Journal: Applied Economics* 1: 34–68.

Hoogendoorn, S. and Van Praag, M. (2012). Ethnic diversity and team performance: A field experiment. Tech. rep., inbergen Institute Discussion Paper No. 12-068/3.

Hunt, J. (forthcoming). The impact of immigration on the educational attainment of natives. *Journal of Human Ressources* .

Isphording, I. E. and Otten, S. (2014). Linguistic Barriers in the Destination Language Acquisition of Immigrants. *Journal of Economic Behavior and Organization* 105: 30–50.

Jensen, P. and Rasmussen, A. W. (2011). The Effect of Immigrant Concentration in Schools on Native and Immigrant Children's Reading and Math Skills. *Economics of Education Review* 30: 1503–1515.

Kahane, L., Longley, N. and Simmons, R. (2013). The Effects of Coworker Heterogeneity on Firm-Level Output: Assessing the Impacts of Cultural and Language Diversity in the National Hockey League. *The Review of Economics and Statistics* 95: 302–314.

Lazear, E. P. (1999). Culture and Language. *Journal of Political Economy* 107: S95–S126.

Machin, S. and Murphy, R. (2018). Paying Out and Crowding Out? The Globalisation of Higher Education. *Journal of Economic Geography. forthcoming* .

Maestri, V. (2017). Can Ethnic Diversity Have a Positive Effect on School Achievement? *Education Economics* 25: 290–303.

McHenry, P. (2015). Immigration and the Human Capital of Natives. *Journal of Human Resources* 50: 34–71.

Migration Advisory Committee (2018). Impact of international students in the UK. Tech. rep.

Ohinata, A. and Ours, J. C. van (2013). How Immigrant Children Affect the Academic Achievement of Native Dutch Children. *Economic Journal* 123: 308–331.

Oster, E. (2017). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics* .

Ottaviano, G. I. and Peri, G. (2006). The economic value of cultural diversity: evidence from us cities. *Journal of Economic geography* 6: 9–44.

Putnam, R. D. (2007). E pluribus unum: Diversity and community in the twenty-first century the 2006 johan skytte prize lecture. *Scandinavian political studies* 30: 137–174.

Top-Universities, Q. (2018). World university ranking 2018 - economics / econometrics.

Trax, M., Brunow, S. and Suedekum, J. (2015). Cultural Diversity and Plant-level Productivity. *Regional Science and Urban Economics* 53.

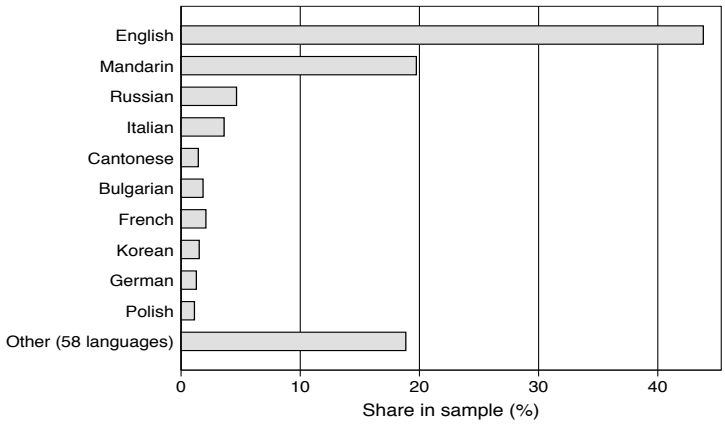
UNESCO (2018). Outbound internationally mobile students by host region.

Wichmann, S., Holman, E. W. and Brown, C. H. (2018). The ASJP Database. *version 18* .

Wooldridge, J. M. (2007). Inverse probability weighted estimation for general missing data problems. *Journal of Econometrics* 141: 1281–1301.

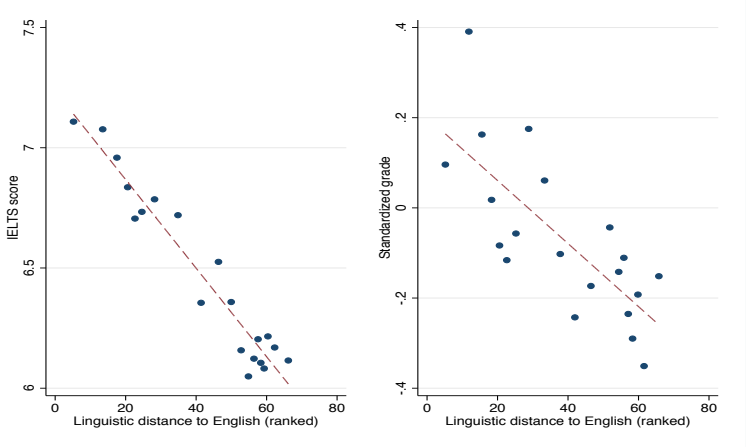
6 Tables & Figures

Figure 1: Sample composition by language background



Notes: This figure displays the share of language groups in the individual sample (n=2,120). Languages are assigned by nationality: each student is assigned the predominant language of the country that the student reports as his/her nationality.

Figure 2: Language skills and grades by distance to English



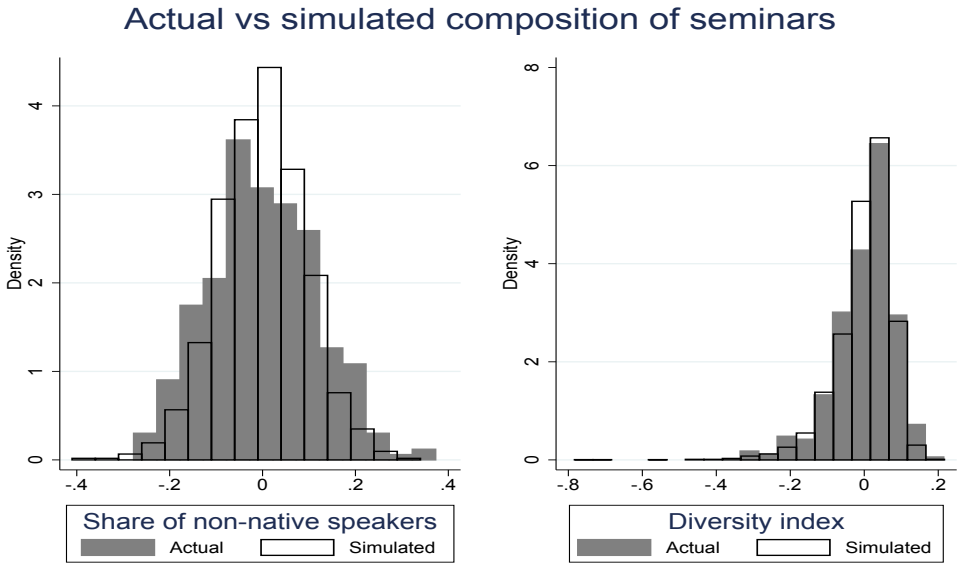
Notes: This figure displays bin scatter plots (20 bins) of the rank in linguistic distance to English to scores achieved in the *International English Language Testing System (IELTS)* (left panel, n=1,949) and the standardised course grade (right panel, n=4,634, non-English speakers only). Results displayed conditional on age, gender, linguistic distance and course×year fixed effects.

Table 1: Sample descriptives

	Sample: Total				Sample: English	Sample: Non-English
A. Dependent variables						
	Mean	*SD*	*Min*	*Max*	*Mean*	*Mean*
Performance						
Std. course grade	0.00	1.00	-4.75	3.14	0.08	-0.07
Course failed	0.17	0.37	0.00	1.00	0.15	0.19
With honors (above 60%)	0.40	0.49	0.00	1.00	0.43	0.38
No. of obs	*8505*				*3909*	*4596*
Educational choices						
Std. average grade in third year	0.00	1.00	-5.40	2.88	0.12	-0.12
Share of numerical courses in third year choices	0.53	0.26	0.00	1.00	0.49	0.58
Share of non-English in third year courses	0.64	0.10	0.08	0.95	0.61	0.66
Difficulty of the courses chosen in third year	0.15	0.07	0.00	0.46	0.14	0.15
No. of obs	*6839*				*3321*	*3518*
Post-graduate outcomes						
Abroad	0.08	0.27	0.00	1.00	0.03	0.16
No. of obs	*2539*				*1582*	*957*
B. Individual characteristics						
	Mean	*SD*	*Min*	*Max*	*Mean*	*Mean*
Ling. dist. to English	51.49	47.89	0.00	104.06	0.00	95.28
Student's age	19.79	1.39	17.00	34.00	19.49	20.05
Female student	0.41	0.49	0.00	1.00	0.34	0.47
No. of obs	*8505*				*3909*	*4596*
C. Seminar characteristics						
	Mean	*SD*	*Min*	*Max*		
No. of students	*26.06*	6.83	10.00	45.00		
Share of non-English speakers	*0.55*	0.14	0.21	0.85		
Blau index of diversity	*0.80*	0.09	0.40	0.93		
No. of obs	*332*					

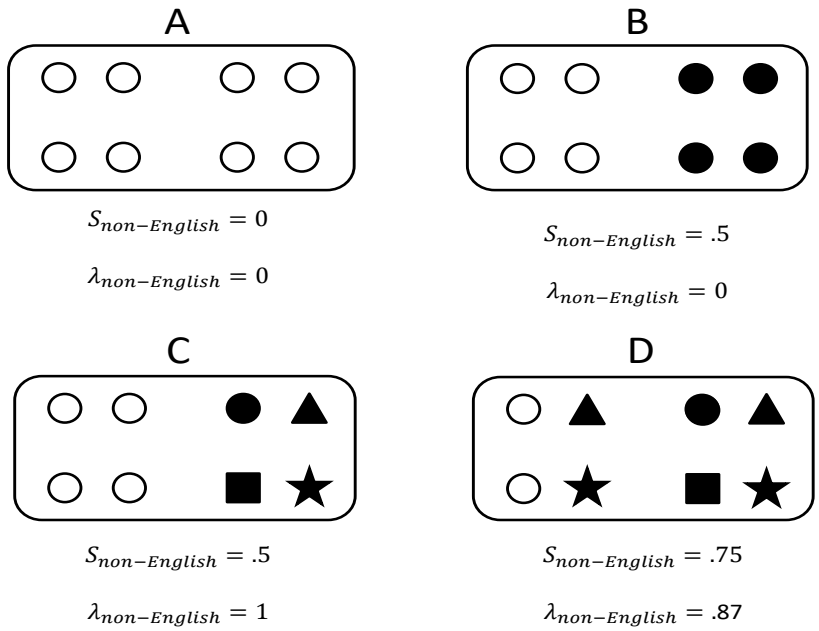
Notes: This table summarises descriptive statistics of individual and seminar characteristics and the dependent variables.

Figure 3: Simulated vs observed seminar composition



Notes: This figure compares observed distributions of the core variables of share of non-English speakers and diversity with simulated distributions based on pure random assignment based on 1,000 replications within courses, holding seminar sizes at observed levels. Variables are displayed as deviations from the $course \times term$ average. Permutation-based p-values of a Wilcoxon rank sum test cannot reject the null of equality between observed and simulated distribution.

Figure 4: Share vs diversity



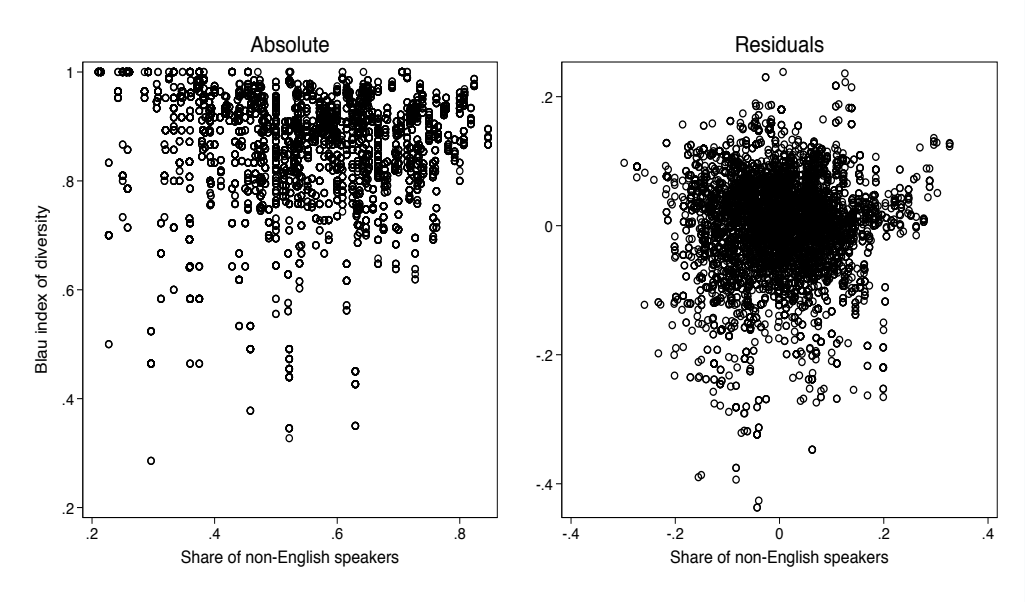
Notes: This figure illustrates the difference between the share of non-English-speaking students and the Blau Diversity Index for classrooms with eight students. Each symbol represents a student. White dots represent English-speaking students while black symbols are for non-English speakers; each shape represents a specific native language.

Table 2: Testing for random assignment

	(1)	(2)	(3)	(4)
Leave-me-out mean/share (seminar)				
Non-English speaker	0.376***	0.022**	0.008	-0.003
	(0.053)	(0.009)	(0.010)	(0.012)
Ling. dist. to English	0.390***	0.025***	0.013	0.001
	(0.052)	(0.009)	(0.010)	(0.012)
first year GPA[a]	0.443***	0.014	0.005	-0.014
	(0.071)	(0.014)	(0.015)	(0.017)
Student's age	0.130**	-0.005	-0.009	-0.010
	(0.055)	(0.010)	(0.010)	(0.011)
Gender: Female	0.211***	0.022**	0.013	0.010
	(0.056)	(0.009)	(0.010)	(0.012)
Language: Mandarin	0.292***	0.024**	0.015	0.014
	(0.056)	(0.010)	(0.011)	(0.013)
Language: Russian	0.236***	0.014	0.014	0.012
	(0.069)	(0.013)	(0.012)	(0.015)
Language: Italian	0.136*	-0.006	-0.015	-0.020**
	(0.072)	(0.009)	(0.009)	(0.010)
Leave-me-out share/mean (urn)	yes	yes	yes	yes
Course × year FE	no	yes	yes	yes
Study program FE	no	no	yes	yes
Day/Time FE	no	no	yes	yes
Seminar leader FE	no	no	no	yes
No. of observations[a]	8505	8505	8505	8505

Notes: This table summarises results of regressions of seminar-wise leave-me-out means/shares on observable student characteristics; each row represents a separate regression. Each regression includes the course/year-wise leave-me-out mean/share and a number of fixed effects. [a]First year GPA is only available for 4,404 observations and does not generally enter our later specifications. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at the individual level, are reported in parentheses.

Figure 5: Variation in share of non-English speakers and diversity



Notes: This figure displays the variation in the share of non-English speakers and the ethnic diversity within the group of non-English speakers, in absolute levels (left panel) and in residuals after applying the within-transformation by course × year, study programme, day × hour, and seminar leader fixed effects (right panel) for the Blau Index. Standard deviations: share of non-English speakers 0.14 (absolute) and 0.09 (residuals), Blau Index 0.10 (absolute) and 0.07 (residuals).

Table 3: Diversity and educational performance

Sample	Grade	Fail	Honour
Total			
Share of non-English	-0.042	0.075*	0.018
	(0.122)	(0.039)	(0.055)
Blau Index	0.210*	-0.113***	-0.036
	(0.128)	(0.042)	(0.058)
Mean of dep. var.	0.000	0.167	0.402
R^2	0.05	0.08	0.13
No. of observations	8505	8505	8505
English			
Share of non-English	0.010	0.061	-0.014
	(0.177)	(0.059)	(0.075)
Blau Index	-0.102	-0.052	-0.029
	(0.157)	(0.063)	(0.076)
Mean of dep. var.	0.082	0.146	0.428
R^2	0.08	0.10	0.15
No. of observations	3909	3909	3909
Non-English			
Share of non-English	-0.036	0.088	0.053
	(0.170)	(0.059)	(0.073)
Blau Index	0.446**	-0.159**	-0.044
	(0.189)	(0.062)	(0.077)
Mean of dep. var.	-0.069	0.186	0.380
R^2	0.08	0.10	0.16
No. of observations	4596	4596	4596
Course × year FE	yes	yes	yes
Study program FE	yes	yes	yes
Day/Time FE	yes	yes	yes
Seminar leader FE	yes	yes	yes
Seminar controls	yes	yes	yes
Individual controls	yes	yes	yes

Notes: This table summarises results of regressions of a set of outcome variables (standardised grade, indicator for failing a course, indicator for receiving an honour (60Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at the seminar level, are reported in parentheses.

Table 4: Mechanisms

Sample	Interaction with English students	Interaction with non-Eng. students	Feeling comfortable using English	Perceived quality of English	Study hours	Own English proficiency
English speakers						
Share of non-natives	-1.143	0.0952	-0.347	-1.868*	2.608	-0.101
	(0.895)	(0.864)	(0.351)	(0.990)	(5.146)	(0.414)
Blau Index	0.534*	-0.205	0.146	0.704	-0.191	0.117
	(0.316)	(0.338)	(0.118)	(0.441)	(2.298)	(0.183)
Mean of dep. var.	0.28	-0.18	0.40	0.19	5.60	0.42
R^2	0.03	0.02	0.02	0.09	0.08	0.02
No. of observations	316	317	316	315	299	317
Non-English speakers						
Share of non-natives	-2.799**	1.947**	-0.958	-0.984	-11.93**	0.672
	(1.132)	(0.843)	(1.333)	(1.034)	(4.830)	(1.084)
Blau Index	1.454**	-1.281***	0.135	0.519	3.242	-0.500
	(0.547)	(0.386)	(0.643)	(0.462)	(2.352)	(0.539)
Mean of dep. var.	-0.40	0.26	-0.57	-0.27	5.82	-0.60
R^2	0.04	0.06	0.02	0.02	0.04	0.02
No. of observations	222	225	223	225	219	224

Notes: This table summarises results of regressions of a set of survey responses on potential mechanisms on the seminar-wise leave-me-out share of non-English speakers and the diversity index. Results by language background (English/Non-English speakers) are derived from split sample models. Individual controls contain age, gender, linguistic distance and whether they are an English speaker or not. Seminar controls are share of females, number of students and mean age. The survey was administered in an adjacent cohort of the autumn semester 2018. The response rate of the survey was 51 percent. Outcomes are standardised from 1-5 scales (columns 1-2: Never to Very Often, column 4: Very uncomfortable to Very comfortable, columns 4 and 6: Very bad to Very good. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at the seminar level, are reported in parentheses.

Table 5: Diversity and third-year choices

	Grade	Share of Numerical courses	Share of Non-English-speaking	Difficulty
Total				
Share of non-English	0.371***	0.014	0.031***	-0.003
	(0.114)	(0.029)	(0.010)	(0.007)
Blau Index	0.088	0.041	-0.013	0.004
	(0.127)	(0.029)	(0.009)	(0.006)
Mean of dep. var.	-0.000	0.533	0.637	0.148
R^2	0.09	0.32	0.47	0.63
No. of observations	6839	6844	6844	6844
English				
Share of non-English	0.301	-0.029	0.005	-0.011
	(0.196)	(0.038)	(0.011)	(0.010)
Blau Index	-0.016	0.005	0.004	-0.012
	(0.173)	(0.034)	(0.011)	(0.010)
R^2	0.10	0.30	0.44	0.63
No. of observations	3321	3326	3326	3326
Non-English				
Share of non-English	0.395**	0.081**	0.058***	0.003
	(0.182)	(0.039)	(0.014)	(0.008)
Blau Index	0.136	0.050	-0.035**	0.011
	(0.185)	(0.044)	(0.016)	(0.008)
R^2	0.11	0.34	0.44	0.65
No. of observations	3518	3518	3518	3518
Course × year FE	yes	yes	yes	yes
Study program FE	yes	yes	yes	yes
Day/Time FE	yes	yes	yes	yes
Seminar leader FE	yes	yes	yes	yes
Seminar controls	yes	yes	yes	yes
Individual controls	yes	yes	yes	yes

Notes: This table summarises results of regressions of a set of outcome variables regarding course choices in third year on the seminar-wise leave-me-out share of non-English speakers and the diversity index. Results by language background (English/Non-English speakers) are derived from split sample models. Individual controls contain age, gender, linguistic distance and they are an English speaker or not. Seminar controls are share of females, number of students and mean age. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at the seminar level, are reported in parentheses.

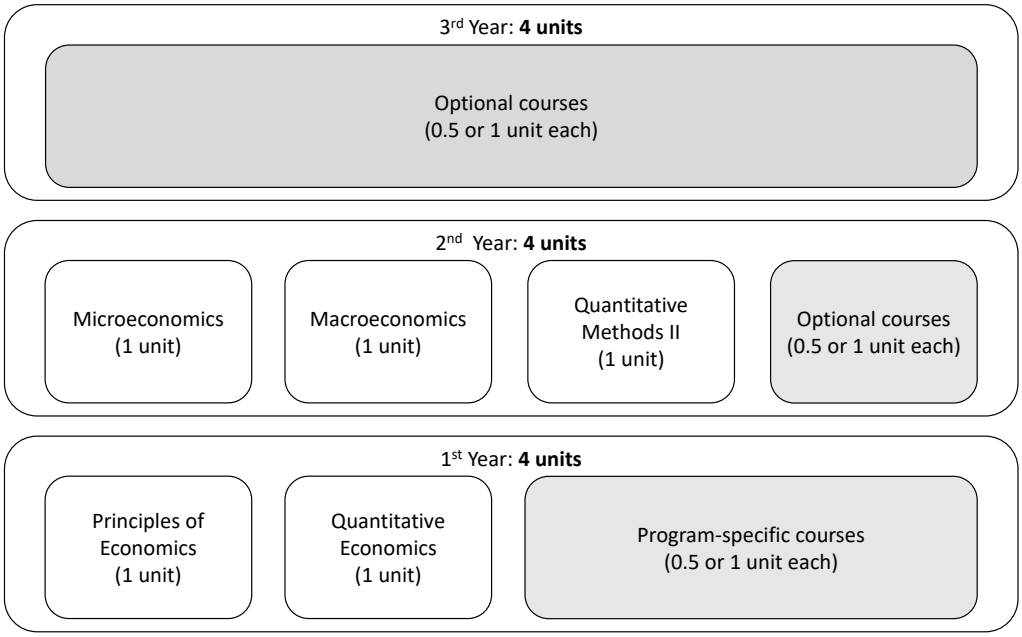
Table 6: Diversity and post-graduation migration

Sample	Response	Abroad
Total		
Share of non-English	-0.048	0.062
	(0.036)	(0.040)
Blau Index	-0.077	-0.089*
	(0.053)	(0.050)
Mean of dep. var.	0.297	0.079
R^2	0.46	0.14
No. of observations	8505	2536
English		
Share of non-English	-0.035	-0.006
	(0.060)	(0.033)
Blau Index	-0.214***	-0.078*
	(0.067)	(0.042)
Mean of dep. var.	0.405	0.027
R^2	0.58	0.12
No. of observations	3909	1580
Non-English		
Share of non-English	0.010	0.343***
	(0.052)	(0.098)
Blau Index	0.029	-0.078
	(0.061)	(0.139)
Mean of dep. var.	0.208	
R^2	0.34	0.17
No. of observations	4596	956
Course × year FE	yes	yes
Study program FE	yes	yes
Day/Time FE	yes	yes
Seminar leader FE	yes	yes
Seminar controls	yes	yes
Individual controls	yes	yes

Notes: This table summarises results of regressions on response rate to a post-graduation survey and to post-graduation migratory decision on the seminar-wise leave-me-out share of non-English speakers and the diversity index. Results by language background (English/Non-English speakers) are derived from split sample models. Individual controls contain age, gender, linguistic distance and whether they are an English speaker or not. Seminar controls are share of females, number of students and mean age. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at the seminar level, are reported in parentheses.

Appendix

Figure A1: Structure of teaching



Notes: This figure describes the teaching structure of the institutional setting. Teaching happens in three consecutive years. Per year, students take four teaching units. In our specifications, we rely on quasi-random assignment into seminars within first- and second-year courses. Non-compulsory optional courses (grey) are not part of our sample. Third-year course choices are regarded as outcomes in Section 4.4.

Table A1: Sample composition by language background

Language	Associated Nationalities	Number of Speakers	Share in sample (%)
ENGLISH	United States Ireland Australia New Zealand Kenya Uganda United Kingdom British Indian Ocean Territory British Overseas Citizen Nigeria Trinidad & Tobago Gambia Canada South Africa	942	44.43
MANDARIN	China Singapore Taiwan	411	19.39
RUSSIAN	Russia Kazakhstan	104	4.91
ITALIAN	Italy	68	3.21
CANTONESE	Hong Kong Macao	39	1.84
BULGARIAN	Bulgaria	39	1.84
FRENCH	France	38	1.79
KOREAN	North Korea South Korea	37	1.75
GERMAN	Germany Austria	32	1.51
POLISH	Poland	25	1.18
ARABIC	Bahrain Saudi Arabia Lebanon United Arab Emirates Libya Oman Morocco Kuwait Egypt Jordan Algeria	24	1.13
GREEK	Greece Cyprus	22	1.04
SWEDISH	Sweden	20	0.94
VIETNAMESE	Vietnam	18	0.85
PORTUGESE	Portugal Brazil Angola	18	0.85
SPANISH	Spain Mexico Columbia El Salvador	18	0.85
AZERBAIJANI	Azerbaijan	17	0.80
HINDI	India	15	0.71
LITHUANIAN	Lithuania	15	0.71
WESTERN PUNJABI	Pakistan	15	0.71
ALL OTHER (48)		203	9.58
Total Sample		2120	

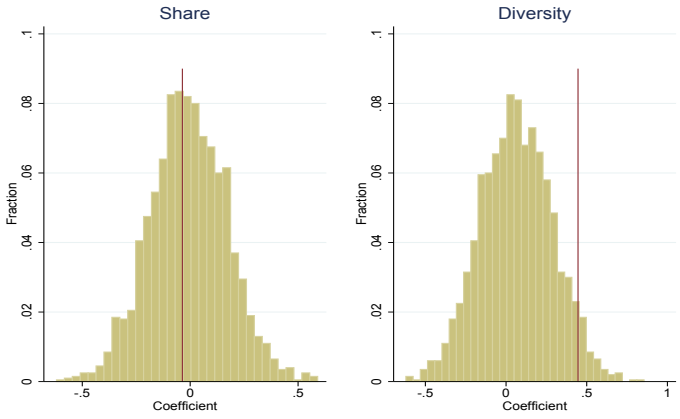
Notes: This table gives the number of individual speakers of top 20 languages as well as the share of that particular language in our full sample.

Table A2: Raw and residual variation in key variables

	Mean	*SD*	*Min*	*Max*
Absolute				
Share of non-English speakers	0.55	0.14	0.21	0.85
Blau index of diversity	0.87	0.10	0.29	1.00
Residualised				
Share of non-English speakers	0.00	0.09	-0.30	0.33
Blau index of diversity	0.00	0.07	-0.44	0.24
No. of obs	*8505*			

Notes: This table shows variation in the share of non-English speakers and the diversity index, in absolute levels and in residualised after controlling for course×year, study programme, day×hour, and seminar leader fixed effects.

Figure A2: Distribution of placebo estimates



Notes: This figure shows the empirical distributions of placebo estimates for the share of non-English-speaking students (left) and ethno-linguistic diversity among those (right) on course grades. The cumulative distribution functions are based on 2000 estimates using a specification similar to the one displayed in column (2) of Table 3 lower panel and using random permutations of seminar ID to compute the treatments. The vertical line indicates the original estimate. Implied p-values are 0.556 (left) and 0.044 (right).

Table A3: Diversity and educational performance, controlling for ability

Sample	Grade	Fail	Honour
Total			
Share of non-English	-0.107	0.093***	-0.003
	(0.089)	(0.034)	(0.045)
Blau Index	0.227***	-0.120***	-0.032
	(0.081)	(0.040)	(0.045)
Mean of dep. var.	0.000	0.167	0.402
R^2	0.48	0.25	0.36
No. of observations	8505	8505	8505
English			
Share of non-English	-0.116	0.086	-0.061
	(0.135)	(0.053)	(0.065)
Blau Index	0.092	-0.091	0.044
	(0.112)	(0.056)	(0.067)
Mean of dep. var.	0.082	0.146	0.428
R^2	0.46	0.23	0.37
No. of observations	3909	3909	3909
Non-English			
Share of non-English	-0.055	0.101*	0.048
	(0.127)	(0.052)	(0.062)
Blau Index	0.353***	-0.141**	-0.077
	(0.126)	(0.058)	(0.059)
Mean of dep. var.	-0.069	0.186	0.380
R^2	0.52	0.28	0.38
No. of observations	4596	4596	4596
Course × year FE	yes	yes	yes
Study program FE	yes	yes	yes
Day/Time FE	yes	yes	yes
Seminar leader FE	yes	yes	yes
Seminar controls	yes	yes	yes
Individual controls	yes	yes	yes
Ability control	yes	yes	yes

Notes: This table summarises results of regressions of a set of outcome variables (standardised grade, indicator for failing a course, indicator for receiving an honour (60Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at the seminar level, are reported in parentheses.

Table A4: Diversity and third-year choices, controlling for ability

	Grade	Share of Numerical courses	Share of Non-English-speaking	Difficulty
Total				
Share of non-English	0.250***	-0.003	0.030***	-0.003
	(0.093)	(0.027)	(0.010)	(0.007)
Blau Index	0.132	0.047*	-0.013	0.004
	(0.103)	(0.028)	(0.009)	(0.006)
Mean of dep. var.	-0.000	0.533	0.637	0.148
R^2	0.40	0.39	0.47	0.63
No. of observations	6839	6844	6844	6844
English				
Share of non-English	0.158	-0.050	0.004	-0.011
	(0.134)	(0.037)	(0.011)	(0.010)
Blau Index	0.170	0.030	0.006	-0.011
	(0.131)	(0.034)	(0.012)	(0.010)
R^2	0.35	0.36	0.44	0.63
No. of observations	3321	3326	3326	3326
Non-English				
Share of non-English	0.301**	0.071**	0.057***	0.003
	(0.138)	(0.033)	(0.014)	(0.008)
Blau Index	0.079	0.044	-0.036**	0.011
	(0.146)	(0.039)	(0.016)	(0.008)
R^2	0.47	0.41	0.45	0.66
No. of observations	3518	3518	3518	3518
Course × year FE	yes	yes	yes	yes
Study program FE	yes	yes	yes	yes
Day/Time FE	yes	yes	yes	yes
Seminar leader FE	yes	yes	yes	yes
Seminar controls	yes	yes	yes	yes
Individual controls	yes	yes	yes	yes
Ability control	yes	yes	yes	yes

Notes: This table summarises results of regressions of a set of outcome variables regarding course choices in third year on the seminar-wise leave-me-out share of non-English speakers and the diversity index. Results by language background (English/Non-English speakers) are derived from split sample models. Individual controls contain age, gender, linguistic distance and whether they are an English speaker or not. Seminar controls are share of females, number of students and mean age. In addition here we control for students' and their peers' GPA. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at the seminar level, are reported in parentheses.

Table A5: Robustness: different diversity specifications

Sample	Baseline	Predominant languages	Nationality
Total			
Share of non-English	-0.042	-0.106	-0.110
	(0.167)	(0.171)	(0.171)
Blau Index	0.210*	0.160	0.244
	(0.126)	(0.127)	(0.149)
Mean of dep. var.	0.000	0.000	0.000
R^2	0.05	0.05	0.05
No. of observations	8505	8505	8505
English			
Share of non-English	0.010	-0.116	-0.135
	(0.187)	(0.190)	(0.191)
Blau Index	-0.102	-0.049	-0.019
	(0.172)	(0.188)	(0.201)
Mean of dep. var.	0.082	0.086	0.103
R^2	0.08	0.08	0.08
No. of observations	3909	3847	3649
Non-English			
Share of non-English	-0.036	-0.062	-0.049
	(0.212)	(0.212)	(0.213)
Blau Index	0.446**	0.332*	0.430
	(0.209)	(0.196)	(0.222)
Mean of dep. var.	-0.069	-0.071	-0.078
R^2	0.08	0.08	0.08
No. of observations	4596	4658	4856
Course × year FE	yes	yes	yes
Study program FE	yes	yes	yes
Day/Time FE	yes	yes	yes
Seminar leader FE	yes	yes	yes
Seminar controls	yes	yes	yes
Individual controls	yes	yes	yes

Notes: This table summarises results of regressions of standardised grades on the seminar-wise leave-me-out share of non-native speakers and different definitions of the diversity index. Individual controls contain age, gender, linguistic distance and whether they are an English speaker or not. In column 2, students are given the predominant language of their country, and are not considered native speakers even if English is an official (but not-predominant) language. In column 3, only the UK nationals are considered to be native speakers. Seminar controls are share of females, number of students and mean age. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at the seminar level, are reported in parentheses.

Table A6: Coefficient stability

Sample	without ability controls		with ability controls	
	without Seminar Controls	with Seminar Controls	without Seminar Controls	with Seminar Controls
Total				
Share of non-English	0.027	-0.042	-0.024	-0.107
	(0.109)	(0.122)	(0.081)	(0.089)
Blau Index	0.181	0.210*	0.182**	0.227***
	(0.124)	(0.128)	(0.084)	(0.081)
Mean of dep. var.	0.000	0.000	0.000	0.000
R^2	0.05	0.05	0.48	0.48
No. of observations	8505	8505	8505	8505
English				
Share of non-English	0.178	0.010	0.070	-0.116
	(0.165)	(0.177)	(0.127)	(0.135)
Blau Index	-0.198	-0.102	-0.028	0.092
	(0.153)	(0.157)	(0.117)	(0.112)
Mean of dep. var.	0.082	0.082	0.082	0.082
R^2	0.08	0.08	0.46	0.46
No. of observations	3909	3909	3909	3909
Non-English				
Share of non-English	-0.096	-0.036	-0.074	-0.055
	(0.160)	(0.170)	(0.115)	(0.135)
Blau Index	0.491***	0.446**	0.370***	0.353***
	(0.186)	(0.189)	(0.124)	(0.126)
Mean of dep. var.	-0.069	-0.069	-0.069	-0.069
R^2	0.08	0.08	0.52	0.52
No. of observations	4596	4596	4596	4596
Course × year FE	yes	yes	yes	yes
Study program FE	yes	yes	yes	yes
Day/Time FE	yes	yes	yes	yes
Seminar leader FE	yes	yes	yes	yes
Individual controls	yes	yes	yes	yes
Seminar controls	no	yes	no	yes
Individual ability	no	no	yes	yes
Peer ability	no	no	no	yes

Notes: This table summarises results of regressions of standardised grades on the seminar-wise leave-me-out share of non-native speakers and the diversity index. Results by language background (English/Non-English speakers) are derived from split sample models. Individual controls contain age, gender, linguistic distance and whether they are an English speaker or not. Seminar controls are share of females, number of students and mean age. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at the seminar level, are reported in parentheses.

Table A7: Questions of the field survey

Question
How do you rate your English proficiency? **[Very bad (1) - Very good (5)]**
In discussions with your classmates of this tutorial, how often do you communicate in a language other than English? **[Never (1) - Very often (5)]**
How comfortable do you feel speaking in English in this tutorial? **[Very uncomfortable (1) - Very comfortable (5)]**
For this course how often do you work with: Native English-speaking students **[Never (1) - Very often (5)]**
For this course how often do you work with: Non-native English-speaking students **[Never (1) - Very often (5)]**
How do you rate the level of English in the seminar discussions? **[Very bad (1) - Very good (5)]**
Apart from attending lectures and seminars, how many hours do you spend working for this course in a typical week? **[Open answer]**

Table A8: The role of Mandarin speakers

Dependent Variable	Share of Mandarin Speakers	Grade	Grade	Grade	Grade	Grade
Total						
Share of non-English	0.284***	-0.036	-0.119	-0.041	-0.149	-0.061
	(0.024)	(0.170)	(0.197)	(0.202)	(0.228)	(0.347)
Blau Index	-0.617***	0.446**	0.627**	0.467	0.315	0.423
	(0.028)	(0.189)	(0.311)	(0.313)	(0.418)	(0.596)
R^2	0.86	0.08	0.08	0.08	0.11	0.19
No. of observations	8505	4596	4596	4596	3117	1478

Notes: This table summarises results of different robustness checks examining the role of Chinese Mandarin speakers for the robustness of our results, Specification (1) relates the share of Mandarin speakers in a seminar to the share of non-English speakers and the diversity index. Columns (2) lists the baseline results similar to Table 3, column (2). Column (3) displays results controlling for the number of Chinese students in the seminar. Column (4) displays results controlling for the share of Chinese students. Columns (5) and (6) repeat this specification separately for Chinese and other non-English speakers. Individual controls contain age, gender, linguistic distance and whether they are an English speaker or not. Seminar controls are share of females, number of students and mean age. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at the seminar level, are reported in parentheses.

Table A9: Robustness of inference

Dependent Variable	Baseline	i.i.d.	robust	course/year	year
Total					
Share of non-English	-0.036	-0.036	-0.036	-0.036	-0.036
	(0.170)	(0.180)	(0.176)	(0.212)	(0.294)
Blau Index	0.446**	0.446**	0.446**	0.446**	0.446**
	(0.189)	(0.208)	(0.209)	(0.209)	(0.174)
R^2	0.08	0.08	0.08	0.08	0.08
No. of observations	4596	4596	4596	4596	4596

Notes: This table summarises results of different robustness checks on inference. Specification (1) displays the baseline specification similar to Table 3, column (2) with standard errors clustered at the seminar level. Column (2) lists results assuming i.i.d. error terms. Column (3) lists results based on robust standard errors. Column (4) displays standard errors clustered on the course×year level. Column (5) applies standard errors clustered on the year level. Individual controls contain age, gender, linguistic distance and whether they are an English speaker or not. Seminar controls are share of females, number of students and mean age. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.